Chapter 1

What Do We Gain from Estimating Euler Equations with Higher-Order Approximations?

1.1 Introduction

Understanding consumption behavior has been a relentless pursuit of economic researchers, in which the estimation of key parameters in the consumer's utility function has occupied a central place in recent years. The importance of these parameters lies in their revealing a consumer's preference such as his willingness to engage in intertemporal substitution, his degree of aversion to risk, as well as the strength of his precautionary saving motive, of which are all relevant in policy formation. Yet despite the many empirical efforts devoted to estimating these parameters, the results are frustrating. Estimates of these structural parameters seem to vary over a wide range, and even worse, they sometimes contradict each other or are inconsistent with their theoretical values.

Carroll (2001b) and Ludvigson and Paxson (2001) argue that those empirical works that estimate preference parameters with a log-linearized or second-order approximated version of the consumption Euler equation are subject to significant approximation bias. This may constitute a compelling explanation of the anomalous results that have been accumulated so far. Even so, this linear approximation technique continues to be useful and convenient in the estimation of the parameters, because the noisy consumption data renders a fully-fledged GMM estimation of the consumption Euler equation unreliable. Since this approximation bias results from omitting higher-order terms in the approximation, we therefore investigate the possibility of reducing this bias through higher-order approximations in this study.

This approximation bias is defined in Ludvigson and Paxson (2001) as the divergence between the true and the estimated parameter values from estimating the linear-approximated version of the Euler equation. It results from improperly approximating the Euler equation that is essentially a nonlinear one. Higher-order terms that are not included in the regression can be regarded as the omitted variables in the regression, which is a direct cause of the biased OLS estimates. Moreover, these higher-order terms are essentially endogenous, and are determined simultaneously with the second-order term in the regression. Second-order approximation with the consumption growth variability term instrumented by individual-specific instruments is incapable of yielding consistent estimators, because it is almost impossible to find instruments that are correlated with the second-order moment term while not being correlated with higher-order moments that are regarded as error terms in the regression. Based on these observations, we believe that this approximation bias can be reduced when these omitted higher-order terms are 'plugged back' into the estimation. The main goal of this chapter is therefore to explore the importance of these higher-order moments in reducing the approximation bias.

To our knowledge, the importance of the consumer's higher-order preference has not been investigated or even discussed in the consumption literature. However, these higher-order terms have been found to be very relevant in financial economics, and are able to explain the empirical failure of the classical CAPM. Harvey and Siddique (2000) and Chung, et al. (2005) find that, when the investor's higher-order preferences are taken into consideration in pricing assets, the significance of non-market risk factors such as SMB (the difference between the return on a portfolio of small size stocks and the return on a portfolio of large size stocks) and HML (the difference between the return on a portfolio of high book-to-market value stocks and the return on a portfolio of low book-to-market value stocks) documented by Fama and French (1995) no longer remains. They find that firm size and book-to-market value are correlated with a stock's skewness, kurtosis, as well as higher-order moments of return distributions that are ignored in the CAPM. The reason why these non-market risk factors seem to be priced is then simply that they proxy for higher-order co-movements that one asset is contributing to the portfolio. This is similar to the argument that ignoring precautionary saving results in the excess sensitivity of consumption because the income level delivers information regarding the income variance.

In the field of portfolio selection, too, researchers who have made various efforts to incorporate higher-order moments beyond Markowitz's mean-variance paradigm are able to find portfolios that outperform those that ignore preferences toward higher-order return moments (Kane, 1982; Simaan, 1993; de Athayde and Flores, 2004; Harvey, et al., 2004).

Rational investors accumulate assets so as to implement their optimal consumption plan, through which their expected lifetime utility is maximized. It is therefore conflicting that individuals care about higher-order moments in selecting assets while ignoring them in the consumption decision process. The approximation bias that omitting these higher-order terms might induce is therefore not surprising.

This relevance of an agent's higher-order preference found in the finance literature thus leads us to the belief that this economic significance will translate into the empirical relevance in estimating the consumer's preference parameters. We are therefore interested in the possibility that implementing higher-order approximations to the Euler equation may drive down the approximation bias that has been found so far.

We investigate the economic significance and empirical relevance of higher-order approximations to the Euler equation through simulations. Our strategy is first to set up a consumer optimization model that is generally used in the consumption literature and then to generate consumption data that researchers will have gathered, given that consumers all behave according to the model. To make our results comparable with previous ones, we set up a model that is exactly identical to the one in Ludvigson and Paxson (2001). In addition, for the same reason, the parameters in the model are all set equal to those of Ludvigson and Paxson (2001), in which the parameter values are all calibrated to best fit the data from PSID. To generate consumption data, the optimization problem is solved numerically for consumption functions that relate consumption to the wealth and income state in each year of the consumer's life. Given these policy rules, we then generate income series for a population of consumers, and their consumption series are then calculated from the optimal consumption rules.

Given this simulated data set, we then investigate the effect on parameter estimates when higher-order approximation is implemented. Our higher-order approximation suggests a linear relationship between expected consumption growth and its higher order moments. When we approximate the Euler equation with a higher-order Taylor's expansion, it suggest that the expected consumption growth should be positively correlated with its higher-order even moments, and negatively correlated with the odd ones. Taking the third-order moment as an example, we know that consumers have a preference toward a positively-skewed consumption pattern, because they prefer a higher probability of a positively extreme event over a higher probability of an extreme event in the negative direction. This means that if consumers expect that their future consumption growth is more likely to be negatively-skewed, they are more worried about the large possibility of a negative extreme outcome. They will thus delay consumption until their concern over the negative extreme outcome is resolved. We will thus expect there to be a negative relationship between consumption growth and its third-order moment.

We examine the effect on bias reduction using both OLS and IV estimation in relation to the higher-order approximated Euler equation. Our results indicate that the approximation bias can be significantly reduced when the higher-order moments are introduced into the estimation, but at a cost of efficiency loss. This reduction in efficiency results from the multi-collinearity between higher-order moments, and the sharp decline in explanatory power of the very instrument set we use to instrument the second-order consumption moment.

Our second set of studies therefore concerns with the approximation order that should be used in the estimation. Since there is a trade off between reducing approximation bias and losing efficiency, we use the mean squared error (MSE) to accommodate these two concerns as the criterion for judging the optimal approximation order. We find it to be affected by the degree of curvature in the utility function and the sample size in estimation. Our final set of simulation then investigate the usefulness of the model selection criteria in providing guidance for selecting the approximation order. Specifically, we compare the MSE of the estimates of the parameter of interest with the approximation order selected by the generalized $\bar{R}^2$ suggested by Pesaran and Smith (1994), information criteria such as the AIC and BIC, and the consistent model and moment selection criteria proposed by Andrews and Lu (2001). We find that the performance of the second-order approximated version of the Euler equation can always be improved simply by allowing for the higher-order moments in the consumption regression, with the approximation order selected by these criteria.

This chapter is organized as follows. Section 2 describes the model's setup and its numerical solution. Section 3 summarizes the problems associated with estimating the Euler equation, and Section 4 then introduces our method of higher-order approximation to the consumption Euler equation. Section 5 presents simulation results when higher-order consumption moments are introduced in estimation. Section 6 then discuss the determination of the optimal approximation order, and the criteria for choosing the approximation order. Section 7 then concludes.

1.2 Model Setup and Its Numerical Solution

1.2.1 The Model

To make our study comparable with the previous literature, we start with a simple but general model of the consumption problem. Specifically, our model is set similar to that of Ludvigson and Paxson (2001), so that their results can serve as a benchmark for comparison purposes. Consumers make consumption and saving decisions in each period so as to maximize their expected lifetime utility. With the assumption of time-separable isoelastic utility functions, their decisions can be viewed as solving the following optimization problem:

$$\text{Max } E_t \left[\sum_{j=0}^{T-t} \beta^j \left(\frac{C_{t+j}^{1-\rho}}{1-\rho} \right) \right]$$
$$\text{s.t.} \quad A_{t+j+1} = (1+r)(A_{t+j} + Y_{t+j} - C_{t+j}), \tag{1.1}$$

where E_t represents the expectation conditional on all information available at time t; C_t equals real consumption at time t; $\beta = \frac{1}{1+\delta}$ is the consumer's subjective discount factor with $\beta \in (0,1]$ and δ is the discount rate. Faced with uncertain labor income Y_t, consumers accumulate non-contingent asset A_t, which pays a gross return $1+r$ in each period, as a means of intertemporally substituting their consumption. We assume that consumers are born with zero assets and that they do not have a bequest motive. This means that we are setting $A_1 = 0$ and $A_{T+1} = 0$.

In the CRRA (constant relative risk aversion) utility function setting, ρ governs the coefficient of relative risk aversion, the elasticity of intertemporal substitution (EIS, the reciprocal of the coefficient of relative risk aversion), as well as the coefficient of relative prudence as defined by Kimball (1990).[1] It is therefore not surprising why so much effort has been devoted to estimating this parameter. We add to this line of research the extent to which adding higher-order moments of the Euler equation can help in uncovering true parameter values correctly.

To complete the description of the model, we specify here the income process facing consumers. Since the result of Ludvigson and Paxson (2001) will serve as a benchmark for later comparison, we follow their work by assuming that income growth evolves according to the following MA(1) process:

$$\ln(Y_{t+1}) = \ln(Y_t) + \mu + \epsilon_{t+1} - \phi\epsilon_t, \tag{1.2}$$

where the innovations to income growth, ϵ_t, are assumed to be normally distributed with mean zero and variance σ^2.

In the model, consumers are not restrained from borrowing. However, we do assume that consumers can only borrow up to an amount that they will be able to repay with

[1]The coefficient of relative prudence is defined as $-\frac{U'''(C_t)C_t}{U''(C_t)}$, which equals $1+\rho$ in our setting.

certainty. That is, the maximum amount that they can borrow is set to be the present discounted value of the minimum realization of income over the reminder of the consumer's life. Note that we have ruled out the possibility of zero income in our model, so that consumers do not act as if there were self-imposed borrowing constraints that prevent them from borrowing, as in Carroll's (1997, 2001a, 2001b) buffer-stock saving models.

1.2.2 Numerical Solution to the Model

Since we are interested in investigating the importance of higher-order moments in uncovering the EIS, consumption data that feature the consumption behavior implied by the model are needed for our later Monte Carlo experiment. This is done by first solving numerically for the optimal consumption rules of the model. With these policy rules in hand, the consumption data can then be generated by simulating realizations of the income process and calculating the consumption levels corresponding to them. Solving the model numerically is therefore the very first task in which we will engage here.

It is now a standard procedure to solve this kind of problem following Deaton's (1991) pioneering work. By defining cash on hand as $X_t \equiv A_t + Y_t$, the dynamic budget constraint of (1.1) can be rewritten as

$$X_{t+1} = (1+r)(X_t - C_t) + Y_{t+1}. \tag{1.3}$$

The optimal decision rules of the utility maximization problem (1.1) can then be derived from this transition equation and the associated Euler equation:

$$C_t^{-\rho} = \beta(1+r)\mathrm{E}_t\left(C_{t+1}^{-\rho}\right). \tag{1.4}$$

Following Ludvigson and Paxson (2001), we normalize the model by dividing each variable by period income Y_t. We define the lower-case variables as their upper-case counterparts divided by the income level for the same period, say, $x_t \equiv X_t/Y_t$. Equation (1.3) can be rewritten as

$$x_{t+1} = (1+r)(x_t - c_t)g_{t+1}^{-1} + 1, \tag{1.5}$$

where $g_{t+1} \equiv Y_{t+1}/Y_t$ is the gross growth rate of income in period $t+1$, which equals $e^{\mu+\epsilon_{t+1}-\phi\epsilon_t}$ under the income process we set in the model. In addition, the Euler equation can be converted to

$$c_t^{-\rho} = \beta(1+r)\mathrm{E}_t\left[(c_{t+1}g_{t+1})^{-\rho}\right]. \tag{1.6}$$

The model is then solved using these normalized variables.

The reason why we re-parameterize the model with income is now clear. Rather than solving for the level of consumption as a function of cash on hand and lagged income, $C_t(X_t, Y_{t-1})$, the model can now be solved by writing the ratio of consumption to income as a function of the ratio of cash on hand to income and the innovation to income growth, $c_t(x_t, \epsilon_t)$. This greatly narrows down the state space we are dealing with, since the range of the possible cash on hand to income is smaller than that of cash on hand, and the distribution of ϵ_t can be characterized more easily in our later work.

The state spaces we mentioned earlier are all continuous. To speed up the calculation, we discretize these state variables into grid points. For innovation to income growth, ϵ_t, we perform a ten-point approximation to its distribution, $N(0, \sigma^2)$. This yields $\epsilon_1, \ldots, \epsilon_{10}$, with each point occurring with a probability of 0.1 in every period. For cash on hand to income, 1,000 grid points are used in the discretization. Since it is well known that the concavity of the consumption function is most dominant when available resources are scarce, we construct the individual points using a triple-exponential growth rate within the state space. This triple-exponential grid we are using here then has the densest grids near the lower bound of x_t, which can help us better capture nonlinearity in the consumption function without too many grid points being needed in solving the model.

By inserting (1.5) into (1.6), the Euler equation facing consumers can be expressed as

$$c_t(x_t, \epsilon_t) = \frac{\beta(1+r)}{10}\sum_{i=1}^{10}\left\{c_{t+1}\left([1+r][x_t - c_t(x_t, \epsilon_t)]e^{-(\mu+\epsilon_i-\phi\epsilon_t)} + 1, \epsilon_i\right)e^{\mu+\epsilon_i-\phi\epsilon_t}\right\}^{-\rho}. \tag{1.7}$$

This Euler equation is then solved recursively backwards from the terminal period T. In the terminal period, the optimal consumption plan is to consume everything, regardless of the realization of ϵ_T. That is, $c_T^*(x_T, \epsilon_T) = x_T$. Given this solution, we can then solve for the values of c_{T-1} that satisfy the Euler equation corresponding to every possible combination of x_{T-1} and ϵ_{T-1}. A numerically optimal consumption rule $c^*_{T-1}(x_{T-1}, \epsilon_{T-1})$ can thus be constructed by linear interpolation between these points. Again, given $c^*_{T-1}(x_{T-1}, \epsilon_{T-1})$, the same method can be applied to construct $c^*_{T-2}(x_{T-2}, \epsilon_{T-2})$. Working backwards, the consumption functions of earlier periods are then solved recursively. Note that these successive consumption rules will not converge as time recedes. For this reason, distinct consumption rules for the T periods have to be derived explicitly, as opposed to using the converged consumption rule as in Carroll (2001b).

1.3 Problems Associated with the Euler Equation Estimation

The consumer's consumption dynamics are best described by the Euler equation facing them, if they are rationally maximizing their expected life-time utility. It is therefore not surprising that plenty of literature that resorts to the Euler equation for studying various issues related to consumption behavior has been accumulated. Among them is to uncover key parameters in the utility function from estimating the linear approximated consumption Euler equation.

The Euler equation is by nature a nonlinear one. To estimate ρ from equation (1.4) directly, the Generalized Method of Moments methodology (GMM) introduced by Hansen (1982) can be employed. Nevertheless, perfect consumption data is required to yield consistent estimates using GMM. However, as investigated by Shapiro (1984) and Runkle (1991), the consumption data are too often measured with error.[2] The presence of measurement error in consumption data will cause GMM estimates of ρ to be inconsistent and thus inapplicable, as discussed in Browning and Lusardi (1996), Carroll (2001b) and Attanasio and Low (2004). This has led researchers to use the linear approximated version of the Euler equation instead. To do this, a commonly used technique is to implement a

[2]Shapiro (1984) reported that a 92% consumption variation in the PSID is noise; Runkle (1991) reported a corresponding figure of 76%.

second-order approximation of equation (1.4):[3]

$$\mathrm{E}_t\left(\frac{C_{t+1}-C_t}{C_t}\right)=\frac{\beta(1+r)-1}{\rho}+\left(\frac{1+\rho}{2}\right)\mathrm{E}_t\left[\left(\frac{C_{t+1}-C_t}{C_t}\right)^2\right]+\nu_t. \qquad (1.8)$$

The above equation is now linear in its parameters and the measurement error problem can hence be readily dealt with.

Equation (1.8) relates the consumption growth rate to uncertainty (as measured by the expected squared future consumption growth, the second-order moment), and the importance of precautionary saving can thus be investigated. The coefficient of this second-order term measures the extent to which prudent consumers defer their consumption in response to future uncertainty. Intuitively, when facing greater uncertainty, consumers postpone consumption to buffer the dramatic consumption variation that this greater uncertainty might cause. The precautionary saving motive thus makes the expected consumption growth rate positively correlated with this consumption risk term. The importance of the precautionary saving motive may then be inferred from the significance of the regression coefficient of this second-order term.

Precautionary saving is defined as the saving behavior consumers engage in to buffer the future uncertainty facing them. It is therefore important in consumption decision-making and wealth accumulating behavior, at least intuitively and theoretically. However, the empirical investigation regarding the importance of precautionary saving has produced mixed results. When investigating the effect of precautionary saving on wealth accumulation, Kazarosian (1997), Lusardi (1997), and Carroll and Samwick (1998) all report that it does explain a significant fraction of wealth accumulation. On the other hand, Kuehlwein (1991), Dynan (1993), and Merrigan and Normandin (1996) find no evidence that the consumption growth rate is affected by consumption uncertainty, indicating that the precautionary saving motive is small or even nonexistent. This is contradictary because it is perceived that individuals accumulate wealth in order to meet their consumption needs, or, to implement their optimal consumption plan. It is therefore puzzling that precau-

[3]Others include the log-linearization of the Euler equation, which is equivalent to a first-order approximation of the Euler equation. Alternatively, it is simply assumed that the consumption growth rate is log-normally distributed, so that the Euler equation can be written as a linear function of the first two moments of the Euler equation.

tionary saving is important in wealth accumulation, while it has little influence in altering the consumption decision.

One compelling explanation for these anomalous results is that studies based on linear approximations to the consumption Euler equation are subject to severe approximation bias in the estimation. Ludvigson and Paxson (2001) and Carroll (2001b) point out that higher-order moments that are regarded as error terms in the linearized Euler equation estimation are essentially *endogenous*, and may be strongly correlated with the regressor (the second-order moment). Omitting these higher terms thus makes the OLS parameter estimates that represent the magnitude of precautionary saving inconsistent, and lead to incorrect inferences. Specifically, in the second-order approximated Euler equation (1.8), the error term ν_t now contains the conditional higher moments of consumption growth. They are endogenously correlated with the second-order term. The orthogonality condition between the regressor and the error term is thus violated. This then leads to biased estimates of the parameters of interest.

Instrumental variable estimation reduces this endogeneity problem, but does not solve it. The omitted higher-order terms are endogenous in the sense that they are jointly determined with the second-order term in the Euler equation, and are thus destined to be correlated with each other. It is therefore extremely difficult to find valid instruments that can well explain the second-order term that is included in the regression of equation (1.8), and at the same time are not correlated with the higher-order moments that are excluded as the error term.

Ludvigson and Paxson (2001) and Carroll (2001b) support these arguments with Monte Carlo experiments. In Ludvigson and Paxson (2001), ρ is set equal to 3 and they generate artificial consumption data that prevails when consumers behave exactly according to the model. They then estimate the second-order approximated Euler equation using OLS and instrumental variable estimation. The results from their Monte Carlo simulation reveal that the true relation between uncertainty and the expected income growth departs substantially from that implied by the linearized regression estimates. The implied value of ρ from the OLS estimation is 0.376, which is only 12.5% of the true value. For instrumental variable estimation, they use education and occupation dummies as instruments. The implied estimate of ρ is 1.816, which is 60% of its true value. The use of instrumental variable estimation corrects some, but not all of, the approximation bias in their experiments.

Carroll (2001b), with a somewhat different model setup, was also able to achieve similar results. In contrast to Ludvigson and Paxson (2001), he constructed a buffer-stock saving model and managed to allow variations in interest rates among different individuals. Yet the results show that log-linearization as well as second-order Taylor expansion are not sufficient to capture the nonlinearities of the Euler equation. He thus concluded that the estimation of the linearized Euler equation should be abandoned because it cannot provide useful information on the structural parameters of interest.

1.4 Higher-Order Approximations to the Consumption Euler Equation

1.4.1 Preferences for Higher-Order Moments

From the previous discussion, we know that the anomalies in investigating the importance of precautionary saving result from ignoring higher-order terms in the Euler equation. We want to stress here, however, that underlying this empirical significance is the economic relevance of these higher-order moments in consumption decision making. That is, the consumer's attitude toward higher moments alters consumption behavior in a significant way. Intuitively, if these terms were irrelevant, although they are still endogenous and correlated with the second-order term, the bias that omitting them might introduce would not be so significant.

For a well-defined utility function $U(\cdot)$, non-satiation and risk aversion suggest that $U'(\cdot) > 0$ and $U''(\cdot) < 0$. In addition, as shown in Arditti (1967), with decreasing absolute risk aversion, we should expect that $U'''(\cdot) > 0$, and that consumers have a preference toward a positive skewed consumption pattern. That is, other things being equal, consumers prefer a high probability of a positive extreme event to a high probability of an extreme event in the negative direction. The preference for a positive third moment has been well established in the finance literature (such as Levy and Sarnat 1984; Sortino and Price, 1994; and Sortino and Forsey, 1996). This has led Kraus and Litzenberger (1976), Lim (1989), and Harvey and Siddique (2000) to incorporate skewness preference into asset pricing models. Their finding is that investors may be willing to sacrifice some expected return if the security would increase the skewness of their portfolio.

The investors' aversion toward negatively skewed asset returns results from their unwillingness to experience a large reduction in wealth, which could dramatically drive down their consumption level. This reveals that people do care about the skewness of their consumption patterns. Despite the fact that the concern over the skewness of future consumption is well established in the finance literature, little attention has been paid to investigating its role in consumption behavior. Approximating the Euler equation up to a second-order moment reveals the researchers' ignorance regarding the consumers' concern over the skewness of consumption. If we perceive that it is important in determining the consumption decision, there is no reason to omit this third-order term when estimating parameter values. The significant approximation bias that omitting this term might induce is therefore not surprising.

$U''''(\cdot)$ reveals the consumer's preference over the kurtosis of the consumption distribution. For utility functions that exhibit *standard risk aversion* as defined by Kimball (1993), Dittmar (2002) shows that $U''''(\cdot) < 0$, and has successfully incorporated this fourth-order moment into the asset pricing framework. Kurtosis describes the extent to which a distribution is weighted toward its tails. A negative fourth derivative of the utility function reveals aversion toward frequent extreme outcomes. For moments greater than the fourth one, Scott and Horvath (1980) prove that, for a well-defined utility function, investors should have positive preferences for odd moments and negative preferences for even ones. Chung, et al. (2005) thus include the 3rd through 10th moments into the asset pricing framework, and show that non-market factors such as book-to-market value and firm size, that are found to be statistically significant in explaining equity returns, may results from ignoring these higher-order terms. Turning to the consumption literature, this finding is somewhat similar to that of Gourinchas and Parker (2002), who argue that ignoring the precautionary saving motive (the second-order moment) can account for the statistical significance of demographic variables such as family size in the first-order or log-linearized Euler equation. Nevertheless, the effect of introducing even higher terms has not been explored in consumption research. We fill this gap by investigating the role that these higher-order terms play in reducing approximation bias when estimating the linear-approximated consumption Euler equation.

1.4.2 How Might Approximation Bias Be Reduced?

From the previous discussion, we know that the insignificance of precautionary saving in consumption behavior might stem from approximation bias induced by improperly approximating the Euler equation. Nevertheless, this approximation technique continues to be useful and convenient in estimating the parameters. Since it is known that the approximation bias arises from omitting the endogenous higher-order terms in the estimation, we try to plug back these omitted terms into the regression and examine their importance in delivering correct parameter values.

The kth order approximation of the consumption Euler equation (1.4) may be expressed as:

$$\begin{aligned} \mathrm{E}_t\left(\frac{C_{t+1}-C_t}{C_t}\right) =& \frac{\beta(1+r)-1}{\rho} \\ &+\sum_{j=2}^{k}(-1)^j\left[\left(\frac{1}{j!}\right)\prod_{z=1}^{j-1}(z+\rho)\right]\mathrm{E}_t\left[\left(\frac{C_{t+1}-C_t}{C_t}\right)^j\right]+\eta_t. \end{aligned} \quad (1.9)$$

The CRRA utility meets the criterion of standard risk aversion defined by Kimball (1993). It therefore has the property that reveals a positive preference for the odd moments, and a negative preference toward the even ones. Equation (1.9) shows the effect of these moments on consumption behavior. Take the third-order moment as an example ($j = 3$). The coefficient of the third-order term is negative ($-[(1+\rho)(2+\rho)/6]$), implying a negative relationship between expected consumption growth and its third-order moment. The intuition underlying this negative correlation is straightforward: if consumers expect their future consumption to be more likely to be positively skewed, they are less worried about a high probability of a negative extreme outcome, and are thus consuming more today which yields a smaller consumption growth rate. Regarding the concern over kurtosis, the coefficient of the fourth-order moment is positive. This means that if consumers perceive that the extreme values are more likely to occur, their aversion toward an extreme outcome deters them from current consumption. Consumption growth thus tends to be greater as a result.

For moments of order greater than the kurtosis, their effect on consumption is less intuitive. Yet we regard these higher-order terms as a means of achieving consistent estimates, just as in the time series regression framework, where *enough* lag terms should be included so that we can obtain consistent estimates of the parameters of interest, although we might not be interested in the coefficients of these lagged terms.

Why might including these higher-order terms reduce approximation bias? Compared with the ν_t term in the second-order approximated equation (1.8), the η_t term in the kth-order approximated Euler equation (1.9) is now much 'cleaner'. Because the 3rd through kth-order moments that cause ν_t to be correlated with the second-order term are now regressors, and the remaining error term η_t is now less correlated with the second-order term. A smaller approximation bias can thus be expected as a result. In the following sections, we then explore the extent to which adding higher terms to the model might reduce approximation bias, as well as whether there is an 'optimal approximation order' in estimating the linearized Euler equation.

1.5 Monte Carlo Evidence

1.5.1 Generating Artificial Data

In this section, we assess the potential gains in reducing approximation bias when higher-order moments are included in the Euler equation estimation by conducting Monte Carlo experiments. To do this, we first generate artificial consumption data that are simulated from groups of individuals who share a common preference and face the same interest rate. These groups of consumers differ only in their income generating processes, which is to say that the only heterogeneity among groups lies in the parameter values of income processes, i.e. μ, ϕ, and σ^2.

As mentioned earlier, we wish to compare our results with those of Ludvigson and Paxson (2001). For this reason, we try to replicate the same consumption data as they generated in their study. They form 16 different educational/occupational groups, with the income process of each group estimated from the PSID. The details of these parameter values that we employ in this research are contained in Table A1 of Appendix A1. Other parameters in the model are assumed to be the same for each cell. Specifically, the coefficient of relative risk aversion ρ is set to be 3 (EIS=1/3), and the interest rate r and discount rate δ are fixed at 0.03 and 0.05, respectively, in our baseline model. Later on, we will change the value of ρ when we want to explore its relationship with the approximation bias and the importance of higher order moments.[4]

The procedure we use to generate simulated data from the model is as follows. First, for each cell, we set T=40 and solve the model for 40 years of consumption functions. We then set up a population of 300,000 consumers who begin their lives with zero cash on hand. Within the population, the number of consumers in each cell is set according to the proportion of each group reported in the PSID. For each consumer in cell j, we draw a random income growth shock ϵ_{1j} from $N(0, \sigma_j^2)$ for his first year of life, where the subscript j refers to the group index, $j = 1, \ldots, 16$. With x_{1j} (=0) and ϵ_{1j} realized, we can use the consumption function for his cell that we just solved to determine his period-one consumption and the saving he carries over to the second year of life. We then draw an income growth shock ϵ_{2j} again, and use equation (1.5) to determine x_{2j}. The consumption function for that period can therefore be used to determine c_{2j}. This exercise is then repeated until we trace out the entire 40-year time path of that consumer. The first and last 5 years of data are discarded, to eliminate the effect of our assumption of initial

[4]Our Monte Carlo study provides no informative information unless the consumption function we solve numerically can well approximate the true consumption function. To check the numerical accuracy of our solutions, accuracy tests proposed by Den Haan and Marcet (1994) are implemented. Since we are using finer grids (1,000 grid points instead of 500) than those of Ludvigson and Paxson (2001), these numerical accuracy tests reveal that the numerical consumption functions that we solved are quite accurate. Interested readers may contact the author for the detailed results.

wealth and the influence of approaching the end of the consumer's life. We therefore now have 300,000 simulated consumers, each with 30 years of consumption data.

Our Monte Carlo experiments then proceed as follows. For each simulation, we draw 1,000 consumers randomly from the 300,000 available in our simulated data set. Consumption data for these 1,000 consumers are then used to perform regressions and estimate the parameter values we are concerned with. This simulation is replicated 300 times to obtain a distribution of parameter estimates and standard error estimates.

1.5.2 Approximation Bias When Only a Second-Order Moment Is Included

We first investigate the severity of ignoring higher-order moments of the Euler equation in estimating ρ. In this baseline model that we are now estimating, we are conducting exactly the same experiment that was conducted in Ludvigson and Paxson (2001). This subsection therefore serves as the reinforcement of their result, and also as a double-check of the resemblance of the simulated data set between the two studies.

For the second-order approximated Euler equation (1.8), we run the following regression:

$$CG_i = \beta_1 + \beta_2 CG2_i + \nu_i,$$

where $CG_i = \left(\frac{1}{T}\right) \sum_{t=1}^{T} CG_{it}$ is the time-averaged consumption growth rate of individual i, and $CG2_i = \left(\frac{1}{T}\right) \sum_{t=1}^{T} CG_{it}^2$ refers to the average squared consumption growth rate. $CG2_i$, the second-order moment in the Euler equation estimation, measures the variability of consumption growth rate of individual i, and is therefore used as a proxy for the uncertainty facing that consumer, which is the driving force behind precautionary saving. When contrasted with equation (1.8), we know that $\beta_2 = \frac{1+\rho}{2}$. Given the parameter values that we set in the model, the regression should yield $\beta_2 = 2$, if we are to correctly uncover the true parameter values using this linearized Euler equation.

The model is estimated using OLS and IV estimation. For the instrumental variable estimation, two sets of instruments are used. The first one uses 16 educational/occupational group dummies as instruments (hereafter IV1), while in the second instrument set (hereafter IV2), 3 education and 6 occupational dummies are used separately, which means that the interactions between these two instruments are excluded.[5]

In addition to estimating β_2, we also perform tests to see whether it differs significantly from its "theoretical value" implied by the linearized Euler equation. Moreover, the Sagan overidentification test (hereafter the OID test) is also conducted to check the validity of the instrumental variables. The results are summarized in Table 1.1. In the table, we report the mean and standard deviation of the estimates of β_2 (whose "theoretical value" is 2) over 300 replications. The average p-value as well as the proportion of rejections obtained in testing $\beta_2 = 2$ and overidentifying restrictions are also reported.

The OLS estimation results are summarized in the first panel of Table 1.1. The estimation of the second-order approximated Euler equation yields estimates of β_2 with an average of 0.641 and a standard deviation of 0.182. These estimates differ significantly from their theoretical value (which is 2 when $\rho = 3$), as there is a 100% rejection rate obtained in testing $\beta_2 = 2$. The downward bias of $\widehat{\beta_2}$ results in an estimate of ρ that is even more remote from its true value. The implied estimate of ρ is 0.281, which is only about one-tenth of the true value of 3.

The IV estimation yields larger estimates of β_2, but they are still biased downwards. Panels 2 and 3 in Table 1.1 present these results. For the IV1 estimation, we first observe that these instruments do explain quite well the variation of $CG2$ in the first-stage regression. The F-statistics for testing the joint insignificance of the instruments are quite large, yielding a 100% rejection rate of the null. The average R^2 of 0.413 also reveals that these group dummies can explain a large proportion of the variability in the consumption growth rate. It should be stressed, however, that this does not imply that these instruments are valid in the sense that they should be orthogonal to the error term in the estimation, since these instruments may as well be correlated with higher-order moments of the Euler equation that are regarded as error terms in the estimation. This

[5] Ludvigson and Paxson (2001) actually use a third instrument set, in which lagged wealth is added to IV2. However, throughout our study, the Monte Carlo results share very similar characteristics to those of IV2. To save space, we thus choose not to report them.

Table 1.1: Second-Order Approximation (ρ = 3,N=1,000)

OLS Estimation				
$\widehat{\beta}_2$	Implied $\widehat{\rho}$	Test for $\beta_2 = 2$	# Rejections	
(std.)	[$\widehat{\rho}/\rho$]	(p-value)	[rej. rate]	
0.641	0.281	-48.31	300	
(0.182)	[0.094]	(0.000)	[1.00]	
IV1 Estimation				
First-Stage Regression				
F-Test	p-value	# Rejections	Rejection Rate	R^2
167	0.000	300	1.00	0.413
Second-Stage Regression				
$\widehat{\beta}_2$	Implied $\widehat{\rho}$	Test for $\beta_2 = 2$	# Rejections	p-OID
(std.)	[$\widehat{\rho}/\rho$]	(p-value)	[rej. rate]	[rej. rate]
1.323	1.646	-11.07	300	0.541
(0.099)	[0.549]	(0.000)	[1.00]	[0.04]
IV2 Estimation				
First-Stage Regression				
F-Test	p-value	# Rejections	Rejection Rate	R^2
309	0.000	300	1.00	0.379
Second-Stage Regression				
$\widehat{\beta}_2$	Implied $\widehat{\rho}$	Test for $\beta_2 = 2$	# Rejections	p-OID
(std.)	[$\widehat{\rho}/\rho$]	(p-value)	[rej. rate]	[rej. rate]
1.328	1.656	-10.51	300	0.4101
(0.101)	[0.552]	(0.000)	[1.00]	[0.07]

Notes. All hypothesis tests are conducted at a 5% significance level. The implied value of β_2 is 2 when $\rho = 3$.

endogeneity problem results in biased estimates of β_2, with a mean of 1.323, which is a far cry from its theoretical value.

The estimation using instrument set IV2 yields similar results. $\widehat{\beta}_2$ is significantly biased downwards, with a 100% rejection rate of the null $\beta_2 = 2$. The implied value of ρ is 1.656, an estimate of only 55.2% of the true value. These results are quite similar to those of Ludvigson and Paxson (2001) and Carroll (2001b). Thus we can draw the same conclusion as theirs: estimation of the *second-order approximated Euler equation* does introduce large systematic bias when estimating the EIS. It therefore seems unwarranted to rely on the second-order linearized Euler equation to reveal the EIS, even for consumers behaving exactly according to the model.

One caveat should be noted before we proceed. It is quite common in the empirical literature to check the validity of instruments using the OID test (such as Dynan, 1993). However, as is shown in the table, this test seems to have little power in detecting violations in overidentifying restrictions implied by the instrument sets we are using. It is clear that the estimates from the IV estimation are remotely biased downwards from their true value, but they are still able to pass the OID tests in most cases. This rejection rate of only up to 7% reveals the fact that the OID tests should be used with caution.[6]

To see the consequences of changing the sample size, we conduct another Monte Carlo experiment with a larger sample size. In each of the 300 replications, we now draw a sample of 10,000 individuals from the data set. The results with this increased sample size are reported in Table 1.2. For the OLS estimation, we first see that the increase in the sample size does not help in reducing the approximation bias. On the contrary, $\widehat{\beta}_2$ even becomes more remotely biased downwards from 0.641 to 0.450, yielding an even more implausible estimate of ρ = -0.1! Yet the information gained from more data available in each estimation does drive down the standard deviation of $\widehat{\beta}_2$ among the 300 estimates. This increase in estimation efficiency along with estimates that are on average more remote from their true value thus makes the test statistics change from -48.31 to -259.4, a symptom that the null of β_2 being equal to its true value is rejected more significantly.

[6] Carroll (2001b) and Ludvigson and Paxson (2001) also make the same caution on the lack of power of the OID tests.

Table 1.2: Second-Order Approximation ($\rho = 3$,N=10,000)

OLS Estimation				
$\widehat{\beta}_2$	Implied $\widehat{\rho}$	Test for $\beta_2 = 2$	# Rejections	
(std.)	$[\widehat{\rho}/\rho]$	(p-value)	[rej. rate]	
0.450	-0.1	-259.4	300	
(0.181)	[-0.033]	(0.000)	[1.00]	

IV1 Estimation				
First-Stage Regression				
F-Test	p-value	# Rejections	Rejection Rate	R^2
1147	0.000	300	1.00	0.277
Second-Stage Regression				
$\widehat{\beta}_2$	Implied $\widehat{\rho}$	Test for $\beta_2 = 2$	# Rejections	p-OID
(std.)	$[\widehat{\rho}/\rho]$	(p-value)	[rej. rate]	[rej. rate]
1.306	1.612	-28.15	300	0.099
(0.063)	[0.537]	(0.000)	[1.00]	[0.78]

IV2 Estimation				
First-Stage Regression				
F-Test	p-value	# Rejections	Rejection Rate	R^2
2163	0.000	300	1.00	0.255
Second-Stage Regression				
$\widehat{\beta}_2$	Implied $\widehat{\rho}$	Test for $\beta_2 = 2$	# Rejections	p-OID
(std.)	$[\widehat{\rho}/\rho]$	(p-value)	[rej. rate]	[rej. rate]
1.309	1.618	-29.92	300	0.064
(0.059)	[0.539]	(0.000)	[1.00]	[0.82]

Notes. All hypothesis tests are conducted at a 5% significance level. The implied value of β_2 is 2 when $\rho = 3$.

Turning now to the IV estimation, for both instrument sets IV1 and IV2, we can estimate β_2 more efficiently from the increased sample size. This can be verified from the decrease in the standard deviation of the 300 replications. Nevertheless, this efficiency gain does not translate into 'correct' estimates, as these estimates seem to center on a point that is now even more distant from the true value. Taking IV1 as an example, even though std.$(\widehat{\beta_2})$ now decreases from 0.099 to 0.063, the average estimate of β_2 decreases as well from 1.323 to 1.306. The result is an average of test statistics that are much more prone to reject $\beta_2 = 2$, and the 100% rejection rate can therefore be expected.

Besides these slight efficiency gains, what draws most of our attention is the dramatic increase in the rejection rate obtained from the OID test. The OID test now rejects overidentifying restrictions with probabilities of 0.78 and 0.82 for IV1 and IV2, respectively. When contrasted with the 0.04 and 0.07 rejection rates when the sample size is 1,000, we see that there is a significant gain in the power of the OID test when the sample size is increased to 10,000. This can clearly be seen from Figure 1.1. In the figure, we plot the empirical distribution of the 300 OID J-statistics, which is asymptotically χ^2 distributed with $l - k$ degrees of freedom under the null of valid instruments (l is the number of instruments, and k is the number of endogeneous variables that are to be instrumented). The top panel in the figure represents the case where IV1 is used as the instrument set, and the bottom panel where we use IV2 as the instrumental variables. The distribution labeled J_{1000} is the distribution of test statistics when the sample size is set to 1,000, and J_{10000} the case when the sample size is 10,000, We also plot the asymptotic distribution of the J-statistic under the null of valid overidentifying restrictions, which is labeled χ^2_{l-k} in the figure. We can see that, for both instrument sets, in spite of the fact that the instruments are not actually exogeneous, the distribution of J_{1000} is quite close to that of χ^2_{l-k}. This reveals the OID test's lack of power when the sample size is 1,000. However, when we enlarge the sample size to 10,000, the distribution of the test statistics, J_{10000}, shifts rightwards in a great deal, and the power of the OID test thus emerges.

So far in this study, we have confirmed Carroll's (2001b) and Ludvigson and Paxson's (2001) finding that the *second-order* approximation of the Euler equation is not sufficient

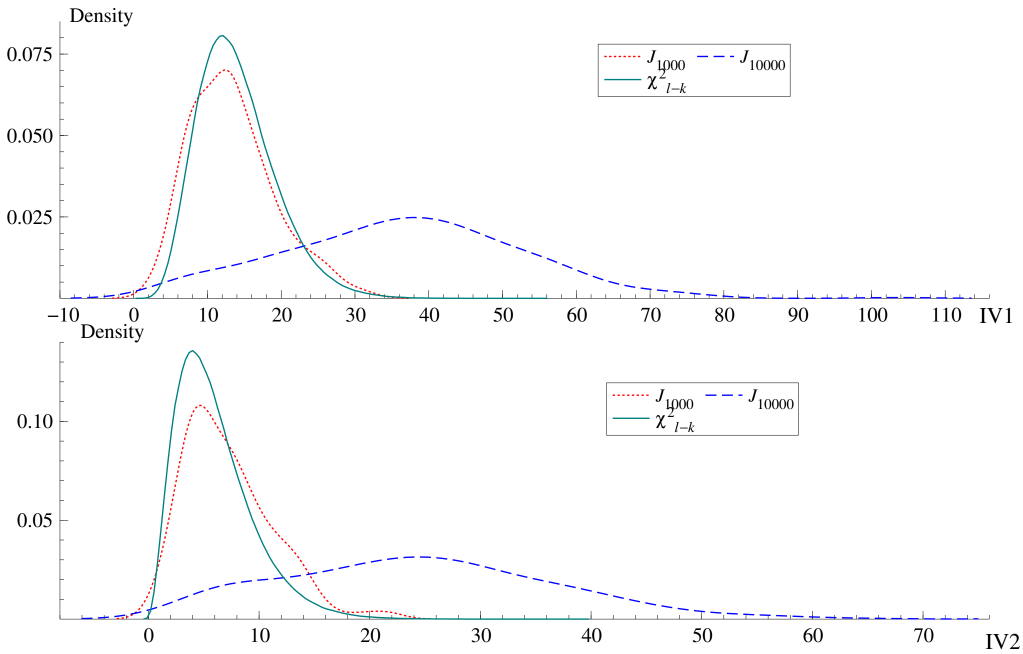

Figure 1.1: OID Test with Different Sample Sizes

to capture the nonlinearities of the problem we are dealing with. Higher-order moments that are regarded as the error terms in regression are thus correlated with independent variables or even instrumental variables. The result is that estimation with the second-order approximated Euler equation provides no convincing information regarding the EIS, even if instrumental variables are used and an enlarged sample size is employed.

1.5.3 Can Approximation Bias Be Reduced When Higher-Order Moments Are Introduced?

As noted above, the large approximation bias in estimating the EIS using a second-order approximated Euler equation results from the endogeneity problem when the omitted higher-order terms are essentially endogenous with respect to the second-order term. Our response to this finding is quite simple and straightforward. Since the bias stems from the fact that higher-order terms omitted in linearization are important and endogeneous, we thus include these higher-order terms in regression by implementing higher-order approximations of the Euler equation. The aim is to better capture nonlinearities coherent with the Euler equation and thus to reduce approximation bias by incorporating these higher-order terms.

In view of the potential gain in the power of the OID test with a larger sample size, for the Monte Carlo simulations that follow, we use a sample size of 10,000 individuals in each replication. For the higher-order approximated Euler equation (1.9), the following regression equation is estimated:

$$CG_i = \beta_1 + \sum_{j=2}^{k} \beta_j CGj_i + \eta_i, \tag{1.10}$$

where k is the approximation order we use in the estimation. As in the case of the second-order approximation, where we use $CG2_i = \left(\frac{1}{T}\right) \sum_{t=1}^{T} CG_{it}^2$ as the proxy for the second-order moment, the j-th order term in equation (1.9) is now proxied by $CGj_i = \left(\frac{1}{T}\right) \sum_{t=1}^{T} CG_{it}^j$.

We investigate the effect of adding higher-order moments to the Euler equation estimation by successively setting k from 2 to 6. Estimations with approximation orders greater than 6 are not conducted, because of the multicollinearity problem that adding these terms might introduce, which will be discussed later in this section. We use both OLS and the IV method in the estimation, as was the case in the previous section and in many empirical studies.

We first report the OLS estimation results in Table 1.3. For each of the approximation orders, the means and standard deviations of the 300 β_2 estimates are reported in the second column of the table. The third column reports the implied value of ρ, calculated from $\widehat{\rho} = 2\widehat{\beta}_2 - 1$, as well as the ratio of this implied $\widehat{\rho}$ to its true value $\rho = 3$. We also test whether β_2 equals its theoretical value (which is 2 when $\rho = 3$), and the results are presented in columns 4 and 5.

As mentioned earlier, estimation with second-order approximation introduces severe approximation bias, with the average estimate of β_2 equal to 0.45 and a 100% rejection rate of $\beta_2 = 2$. When higher-order terms are introduced, however, OLS estimates of β_2 by no means perform better then the second-order approximated ones. There is no evidence revealing that these estimates will be closer to their true value when higher-order moments are added to the regression. Even in the best case where fourth-order approximation is conducted, $\widehat{\beta}_2$ is only 1.109, and the implied estimate of ρ is 1.218, which is only about 41%

Table 1.3: Higher-Order Approximation: OLS Estimation ($\rho = 3$,N=10,000)

Approx. Order	$\widehat{\beta}_2$ (std.)	Implied $\widehat{\rho}$ $[\widehat{\rho}/\rho]$	Test for $\beta_2 = 2$ (p-value)	# Rejections [rej. rate]
2	0.450	-0.1	-259.4	300
	(0.181)	[-0.033]	(0.000)	[1.00]
3	0.879	0.758	-90.97	300
	(0.143)	[0.253]	(0.000)	[1.00]
4	1.109	1.218	-51.82	300
	(0.094)	[0.406]	(0.000)	[1.00]
5	0.876	0.752	-43.38	300
	(0.295)	[0.251]	(0.000)	[1.00]
6	0.252	-0.496	-62.8	300
	(0.370)	[-0.165]	(0.000)	[1.00]

Notes. All hypothesis tests are conducted at a 5% significance level. The implied value of β_2 is 2 when $\rho = 3$.

of its true value. The sixth-order approximation yields the worst results. The $\widehat{\beta}_2$ are now even more remote from their theoretical value, yielding an implausible negative estimate of ρ. What is even worse, as more terms are included in the regression, the variability of the estimates of β_2 is increased as well. The standard deviation of β_2 estimates increases from 0.181 to 0.370, without a significant reduction in approximation bias. The reason why adding higher-order terms reduces estimation efficiency is twofolds. With the same information available from the data, more coefficients that are now estimated as higher terms are included. The estimation efficiency decreases as a result. The second driving force lies in that these higher-order terms are endogenously correlated, because they are all determined simultaneously through the Euler equation in the consumer's decision process. This correlation leads to multi-collinearity problem and therefore drives down the efficiency in the estimation.[7]

[7]The efficiency does not seem to decrease when the Euler equation is approximated up to the 3rd or 4th order, as the standard deviation in estimating β_2 decreases. This might result from the fact that when we implement the 3rd or 4th approximations, the estimates are actually becoming closer to their theoretical values. When 5th- or 6th- order moments are included, however, the efficiency does decrease as the mean estimates are now even further from 2.

The result that adding higher-order terms provides little benefit in reducing approximation bias may appear to be at odds with the intuition that approximation bias comes from these omitted higher-order terms. A direct cause of this inconsistency comes from the fact that even though higher-order terms are included, the expectation error terms in equation (1.9) that are sorted into the error term η_i in the regression (1.10) may still be correlated with independent variables in the regression. This is the very reason why instrumental variable estimation is so often employed in empirical studies. We thus proceed by investigating whether this IV technique with higher-order terms included in the estimation can constitute estimates that reveal true values of the underlying parameters that we are interested in.

The results of using IV1 as the instrument set are summarized in Table 1.4, where we expand the Euler equation up to the sixth-order moment. The second-stage regression results are reported in the top panel of the table, while the first-stage regression results are summarized in the bottom panel. In the first-stage regression, when the Euler equation is linearized to the k-th order, $CG2$ through CGk are all instrumented using group dummies IV1. That is, each moment is regressed on IV1 using OLS estimation, and the fitted value of each moment then serves as the regressor in the second-stage regression. We first observe that when higher-order terms are instrumented with IV1, the explanatory power of these instruments decreases as we move to higher-order terms. This can apparently be seen from the F-statistics and the R^2 reported in the table. For CGk that is to be instrumented, we perform tests to determine whether these instruments are jointly insignificant in explaining CGk. As k increases, the average F-statistic decreases from 1147 to 1.175, and the p-value increases from 0 in $CG2$ to 0.488 in $CG6$. Furthermore, the rejection rate of joint insignificance drops sharply from 100% to 17%, and the R^2 decreases from 0.276 to 0.002 as well. This decrease in explanatory power does have consequences regarding the performance of the second-stage regression, which will be evident in our later discussion.

For the second-stage regression, the first point to be made is that estimates of β_2 that exhibit smaller biases can be achieved by adding higher-order terms to the regression.

Table 1.4: Higher-Order Approximation: IV1 Estimation ($\rho = 3$,N=10,000)

Approx. Order	$\widehat{\beta}_2$ (std.)	Implied $\widehat{\rho}$ [$\widehat{\rho}/\rho$]	Test for $\beta_2 = 2$ (p-value)	# Rejections [rej. rate]	p-OID [rej. rate]
2	1.306	1.612	-28.15	300	0.099
	(0.063)	[0.537]	(0.000)	[1.00]	[0.78]
3	1.557	2.14	-9.01	297	0.109
	(0.093)	[0.713]	(0.002)	[0.99]	[0.78]
4	1.765	2.53	-1.90	115	0.479
	(0.215)	[0.843]	(0.349)	[0.38]	[0.32]
5	1.884	2.768	-0.496	57	0.709
	(0.302)	[0.92]	(0.562)	[0.19]	[0.10]
6	1.899	2.798	-0.572	36	0.863
	(0.422)	[0.933]	(0.611)	[0.12]	[0.05]
First-Stage Regression Results					
Variable	F-Test	p-value	# Rejections	Rejection Rate	R^2
CG2	1147	0	300	1.00	0.276
CG3	24.12	0.087	250	0.83	0.022
CG4	3.892	0.208	189	0.63	0.004
CG5	1.648	0.367	98	0.33	0.002
CG6	1.175	0.488	52	0.17	0.002

Notes. All hypothesis tests are conducted at a 5% significance level. The implied value of β_2 is 2 when $\rho = 3$.

By examining the average estimates of β_2 in the second column, we can see that they do 'converge' to their theoretical value of 2, as the estimate now increases from 1.306 in the second-order approximated case to 1.899 when the Euler equation is approximated to the sixth order. The implied value of ρ now increases from 1.612 to 2.798, yielding an estimate that is 93.3% of its true value. When contrasted with the estimate of ρ that is only 53.7% of its true value, we can see that these higher-order terms do in fact drive down approximation biases by eliminating correlations between the regressors and the error terms.

These less-biased estimates are not achieved without any price being paid. When more higher-order terms are introduced, the standard deviation of the β_2 estimates now increases from 0.063 to 0.422. Three driving forces are responsible for this dramatic reduction in efficiency. The first two are those that we have discussed previously. As more coefficients are estimated and the multicollinearity problem is induced when higher-order moments are included, the estimation efficiency reduces as a result. The third main driving force is that the predicting power of the instrument set drops dramatically when higher-order terms are introduced, the resulting estimates are thus relatively less efficient.

Turning now to the test of $\beta_2 = 2$, it is rejected only at a 12% rejection rate when $k = 6$. This is a drastic reduction when compared with the 100% rejection rate where only a second-order moment is included. This reduction in the rejection rate surely results from the fact that these estimates are now less biased. Nevertheless, the increase in the standard error of $\widehat{\beta_2}$ also broadens the confidence interval. This also gives rise to test statistics that are rejected less frequently.

In the last column of the table, the OID-test results on the plausibility of the instrumental variables are reported. The p-value increases from 0.099 to 0.863, and the rejection rate of the overidentifying restrictions decreases from 78% to 5%, when the approximation order is increased from 2 to 6. This suggests that these instruments are now more 'valid' in the sense that they are much less correlated with the error terms. The result that the β_2 estimates are less biased and can thus better help uncover the true value of ρ is therefore not surprising.

The reason why adding higher-order terms drives down approximation bias is quite clear and intuitive. Approximation bias in estimating second-order approximated Euler equations results from the omitted higher-order terms that are essentially endogenous. It is therefore almost impossible to find an instrument set that can explain the consumption risk while remain correlated with these higher-order terms. This is also evident from our first-stage estimation results. We can see that at least up to $CG4$, the instrumental variables are significant in explaining these higher-order terms in most cases. Omitting these terms means that they are regarded as error terms in the regression. This then leads to violations of the orthogonal condition between the instrumental variables and the error

Table 1.5: Higher-Order Approximation: IV2 Estimation ($\rho = 3$,N=10,000)

Approx. Order	$\widehat{\beta}_2$ (std.)	Implied $\widehat{\rho}$ $[\widehat{\rho}/\rho]$	Test for $\beta_2 = 2$ (p-value)	# Rejections [rej. rate]	p-OID [rej. rate]
2	1.309	1.618	-29.92	300	0.064
	(0.059)	[0.539]	(0.000)	[1.00]	[0.82]
3	1.581	2.162	-6.657	289	0.070
	(0.119)	[0.721]	(0.009)	[0.96]	[0.81]
4	1.989	2.978	-0.856	66	0.553
	(0.443)	[0.993]	(0.511)	[0.22]	[0.23]
5	2.170	3.340	1.12	18	0.745
	(0.847)	[1.11]	(0.617)	[0.06]	[0.05]
6	1.985	2.97	-0.322	10	0.792
	(1.66)	[0.990]	(0.725)	[0.03]	[0.06]
First-Stage Regression Results					
Variable	F-Test	p-value	# Rejections	Rejection Rate	R^2
CG2	2163	0.000	300	1.00	0.255
CG3	46.04	0.030	250	0.83	0.021
CG4	7.10	0.098	214	0.71	0.004
CG5	2.84	0.166	148	0.49	0.002
CG6	1.96	0.227	96	0.32	0.001

Notes. All hypothesis tests are conducted at a 5% significance level. The implied value of β_2 is 2 when $\rho = 3$.

terms. The resulting endogeneity thus makes correctly uncovering the structural parameters unattainable. Retrieving these higher-order terms back as independent variables then 'purges' the error terms by reducing their correlations with the instrumental variables. The elimination of approximation bias can therefore be expected.

What if IV2 is used as the instrument set? The results are presented in Table 1.5. The approximation bias declines sharply as higher-order terms are introduced, especially when the Euler equation is approximated to the fourth-order moment. Specifically, when the skewness and kurtosis are taken into account in the estimation, the approximation bias can be reduced dramatically from 0.691 (2-0.309) to 0.011 (2-1.989). This is quite appealing because it reveals that as long as we take into consideration the consumers' preferences toward skewness and kurtosis, an estimate of ρ that is 99.3% of its true value can now easily be achieved.

It should be noted, however, that when still higher-order terms (such as the fifth- and sixth-order moments) are introduced, the approximation bias cannot be further reduced. While still yielding smaller approximation bias than the second-order approximated Euler equation estimation, the dramatic loss in efficiency cannot be overlooked. When we approximate the Euler equation to the sixth order, the variation in $\widehat{\beta}_2$ becomes incredibly large with a standard deviation of 1.66. The information provided by each estimation is therefore very vague, and the resulting confidence interval is so wide that the null will not be rejected within a broad range of $\widehat{\beta}_2$.

It therefore appears that there is no reason to include these 5th or 6th moments, since there is no reduction in bias, but only a decrease in efficiency. In trading off between reducing approximation bias and losing efficiency, there seems to exist some 'optimal approximation order' that makes the estimates less biased, without too much inefficiency being incurred. We will turn to explore this issue later. Our criterion of choosing this optimal order and its relationship with the preference parameters will also be discussed.

To conclude the discussion of the baseline case, adding higher-order moments to the regression is not useful when the OLS technique is used in the estimation. Nevertheless, these higher-order terms do help reduce approximation bias when IV estimation is employed, even though the cost in terms of reduced efficiency has to be paid. We are thus not pessimistic regarding relying on the linearized version of the Euler equation to uncover the structural parameters, at least under the general parameter setup of Ludvigson and Paxson (2001).

1.6 The Optimal Approximation Order

As mentioned earlier, the approximation bias can be reduced when higher-order terms are introduced, although at the cost of efficiency loss. It is therefore important to find a measure that accommodates both the concern over biasedness and efficiency, in order to decide which approximation order we should use. A criterion that recognizes this possible tradeoff is the mean squared error (MSE). The MSE, as we all know, is the sum of the bias squared and the variance of the estimates. It therefore perfectly fits our concern in determining the optimal approximation order.

In the following analysis, we thus first use the MSE to determine the optimal approximation order, and its relation to the sample size and the value of ρ. After that, knowing that the true value of ρ is rarely known, we investigate the usefulness of the model and moment selection criteria that are often employed in the empirical work. Their performances in selecting the approximation order are then compared with the MSE of the β_2 estimates selected by each criterion.

1.6.1 Sample Size and the Optimal Approximation Order

Here we explore the relation between the sample size and the optimal approximation order. The optimal approximation order is defined here as the expansion order of the consumption Euler equation that yields the minimum MSE of the β_2 estimates. Intuitively, we would expect that β_2 can be estimated more efficiently and the optimal approximation order would be greater as we increase the sample size. The reason is that the increased sample size provides more information in estimating the structural parameters, and that the optimal approximation order is now more 'convergent' to the true order of infinity.

We now present the Monte Carlo results of the different sample sizes and approximation orders in Table1.6.[8] In this benchmark case, we set $\rho = 3$. And to make our results more robust, we increase the replication times up to 1,000. The sample size N is set to be 1000, 2000, 5000, 8000, 10000, 20000, 50000, 80000, and 100000, respectively. Estimation results are reported up to the 10th order.[9]

We first look at the bias of the β_2 estimates. As is also evident from Figure1.2, the

[8] The OLS estimation results are not reported because adding higher-order terms is of no use in reducing approximation bias. Since we are interested in the potential benefit of adding higher-order terms, we thus choose not to report the OLS results. Moreover, as the IV2 estimation results are qualitatively similar to the IV1 results, to save space, we do not report them either.

[9] With the 16 instrumental variables we are using, we can actually approximate the Euler equation up to the 16th order. But the multi-collinearity problem becomes so severe that the variance of β_2 estimates become very huge. We thus choose not to report them in the table.

Table 1.6: β_2 Estimates with Different Sample Sizes and Approximation Orders ($\rho = 3$)

Order	N=1000	N=2000	N=5000	N=8000	N=10000	N=20000	N=50000	N=80000	N=100000
2	1.294	1.296	1.296	1.302	1.303	1.308	1.308	1.309	1.310
	(0.158)	(0.148)	(0.100)	(0.071)	(0.063)	(0.045)	(0.032)	(0.021)	(0.019)
3	1.605	1.584	1.562	1.560	1.552	1.538	1.502	1.483	1.476
	(0.192)	(0.138)	(0.105)	(0.095)	(0.100)	(0.089)	(0.084)	(0.084)	(0.077)
4	1.511	1.662	1.755	1.768	1.768	1.738	1.680	1.674	1.653
	(0.460)	(0.331)	(0.235)	(0.210)	(0.221)	(0.207)	(0.161)	(0.152)	(0.130)
5	1.236	1.495	1.767	1.845	1.865	1.934	1.934	1.913	1.904
	(0.491)	(0.424)	(0.327)	(0.293)	(0.281)	(0.259)	(0.293)	(0.326)	(0.321)
6	1.400	1.476	1.677	1.766	1.892	1.899	2.010*	2.046	2.046
	(0.506)	(0.396)	(0.326)	(0.298)	(0.303)	(0.245)	(0.285)	(0.327)	(0.336)
7	1.701*	1.814*	1.820	1.854	1.852	1.971	1.922	1.978	1.982*
	(0.627)	(0.437)	(0.336)	(0.292)	(0.281)	(0.217)	(0.257)	(0.283)	(0.281)
8	1.164	1.696	2.019*	2.036*	2.025*	1.985*	1.960	1.975	1.912
	(0.969)	(0.883)	(0.549)	(0.394)	(0.438)	(0.274)	(0.184)	(0.207)	(0.217)
9	1.199	1.139	1.587	1.892	2.126	2.037	1.970	1.980*	1.980
	(5.780)	(0.903)	(1.907)	(0.804)	(1.114)	(0.848)	(0.348)	(0.637)	(0.200)
10	1.087	1.316	1.329	1.596	1.564	2.150	1.923	1.963	1.878
	(4.099)	(1.610)	(5.019)	(3.845)	(1.390)	(2.544)	(0.889)	(1.430)	(1.597)

Notes. Figures reported in the table are the average estimates of β_2 among 1,000 replications. Standard errors in parentheses. β_2 should be equal to 2 when $\rho = 3$.

* The approximation order that yields the minimum absolute value of bias for each sample size.

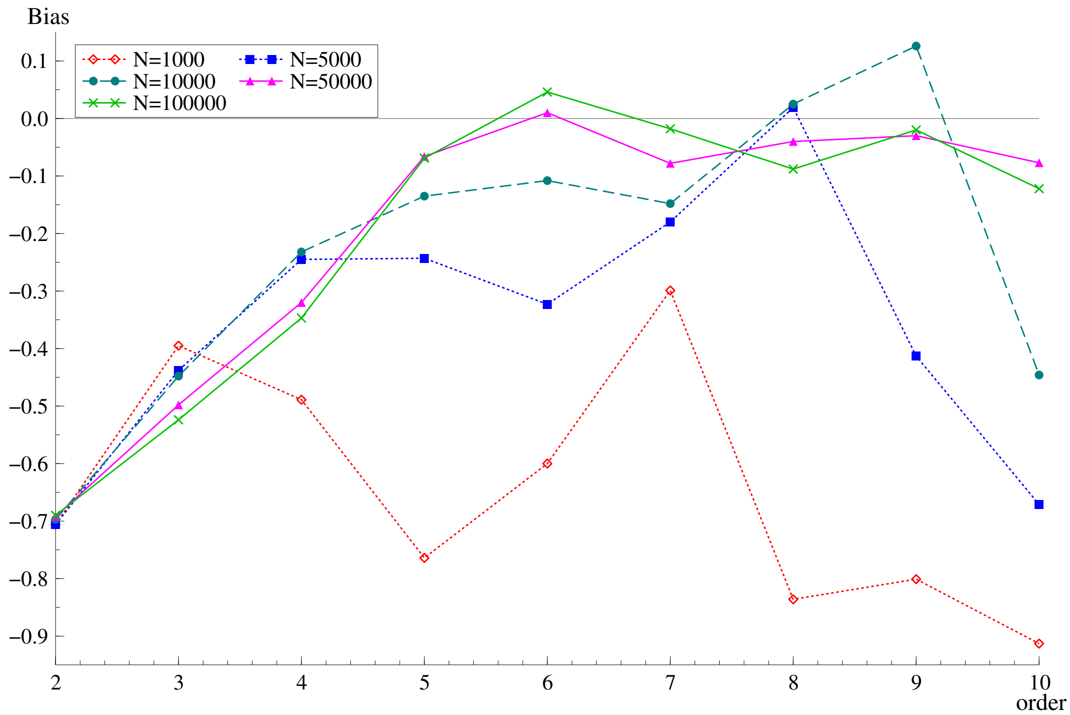

Figure 1.2: Biases of Different Sample Sizes and Approximation Orders

approximation bias can be significantly reduced if the *proper* approximation order is used, except for the case where N=1000. The approximation bias seems to be like a 'inverse U' shaped curve in the figure. This means that the bias declines first as the higher-order consumption moments are added into the regression. Nevertheless, when the bias reaches its minimum which is fairly close to zero, the bias then increases dramatically as even more higher terms are included. The former results from that the error term are now more 'purged' and are therefore much more orthogonal to the regressors when higher-order terms are introduced. The later then results from that as more and more higher-order moments are included into the model, few further information on the consumption behavior (distribution) can be provided. The resulting multi-collinearity thus leads very inaccurate β_2 estimates.

We also find that the increase in sample size does not help in reducing the approximation bias of the 2nd-order approximated estimations. They consistently yield estimates that are close to 1.300, which implies a bias about -0.7. The advantage of increasing the sample size is quite obvious, however, when the higher-order moments are introduced into

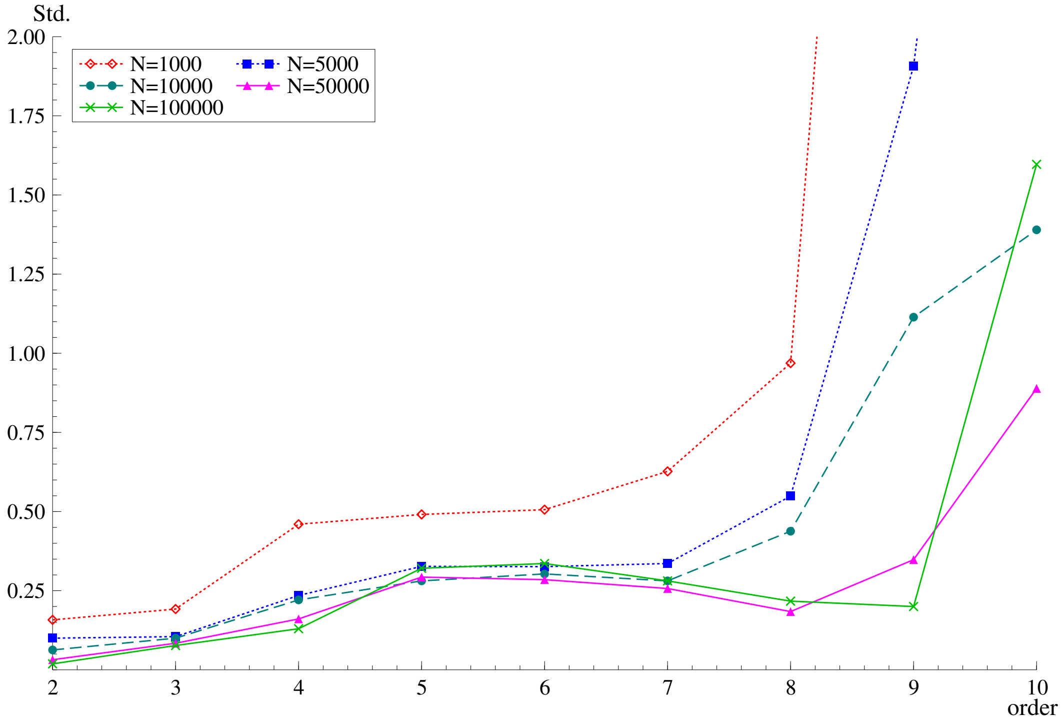

Figure 1.3: Standard Deviation of the β_2 Estimates from Different Sample Sizes and Approximation Orders

the model. The minimum of the bias (in absolute value) can be reduced from 0.299 when N=1000 (7th-order approximation) to 0.025 when N=10000 (8th-order approximation) and to 0.018 (7th-order approximation) when N=100000.

The standard error of the β_2 estimates are depicted in Figure 1.3. Apart from what we have mentioned before that there is a great reduction in estimation efficiency when higher-order consumption moments are introduced into the model, the relation between the estimation efficiency and the sample size is also obvious. For most approximation orders, the estimation efficiency can be enhanced when the sample size is enlarged. The reason is that the greater sample size provides more information of the parameter of interest, and the parameter can thus be estimated more precisely.

We now turn to the MSE criterion that accommodates both the concern over the unbiasedness and the estimation efficiency. The MSEs of the β_2 estimates of different approximation order and sample sizes are summarized in Table 1.7. For each sample size, the 'optimal approximation orders' that yields the minimum MSE are labeled with '$*$' in

Table 1.7: MSE of β_2 Estimates ($\rho = 3$)

Order	N=1000	N=2000	N=5000	N=8000	N=10000	N=20000	N=50000	N=80000	N=100000
2	0.524	0.517	0.505	0.492	0.490	0.481	0.482	0.477	0.477
3	0.193*	0.192*	0.203	0.203	0.211	0.222	0.255	0.274	0.280
4	0.452	0.224	0.116*	0.097*	0.103	0.119	0.129	0.130	0.138
5	0.824	0.436	0.161	0.110	0.097*	0.071	0.091	0.114	0.113
6	0.616	0.432	0.210	0.143	0.103	0.070	0.081	0.109	0.115
7	0.483	0.225	0.146	0.106	0.109	0.054*	0.072	0.081	0.079
8	1.638	0.872	0.301	0.157	0.193	0.075	0.038*	0.044*	0.051
9	34.053	1.557	3.808	0.659	1.255	0.720	0.122	0.406	0.040*
10	17.634	3.060	25.638	14.950	2.121	6.495	0.796	2.047	2.566

Notes. Replication for 1,000 times.

* The 'optimal approximation order' with the minimum MSE.

the table. The result is that when we increase the sample size, the optimal approximation order increases. A smaller MSE can be achieved as well. The minimum of the MSE can be reduced from 0.193 when N=1000 to 0.097 when N=10000, and to 0.040 when N=100000.

This pattern can be seen clearly in Figure 1.4. The MSEs seems to exhibit a U-shaped pattern when we increase the approximation order. This means that there does exist some approximation order that yields the minimum MSE. The idea that we do not suggest unlimited expansion of the consumption Euler equation can thus be verified. We also find that with a greater sample size, the U-shaped MSE curve shifts rightwards and downwards, implying a smaller minimum MSE achieved with a greater approximation order.

To conclude the subsection, we depict the optimal approximation order and the corresponding MSE of the different sample sizes in Figure 1.5. As is evident from the figure, the optimal approximation order increases when a greater sample size is used. The sharp decline in MSE reveals the fact that a smaller bias can be achieved and the efficiency gain from increased sample size.

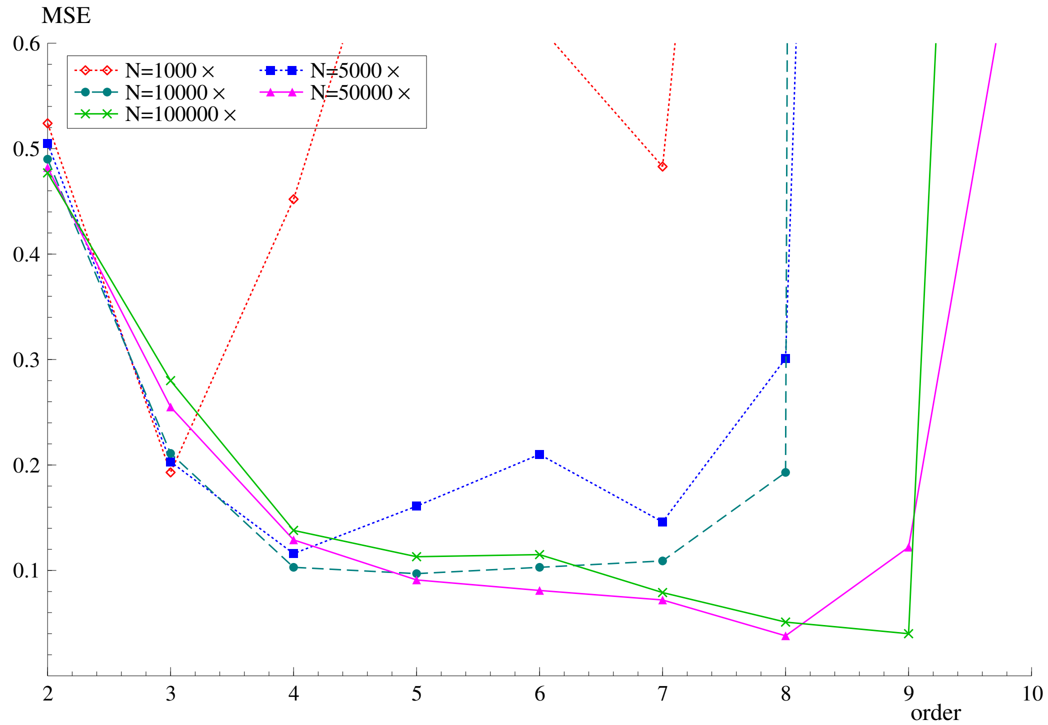

Figure 1.4: MSE with Different Sample Sizes and Approximation Orders

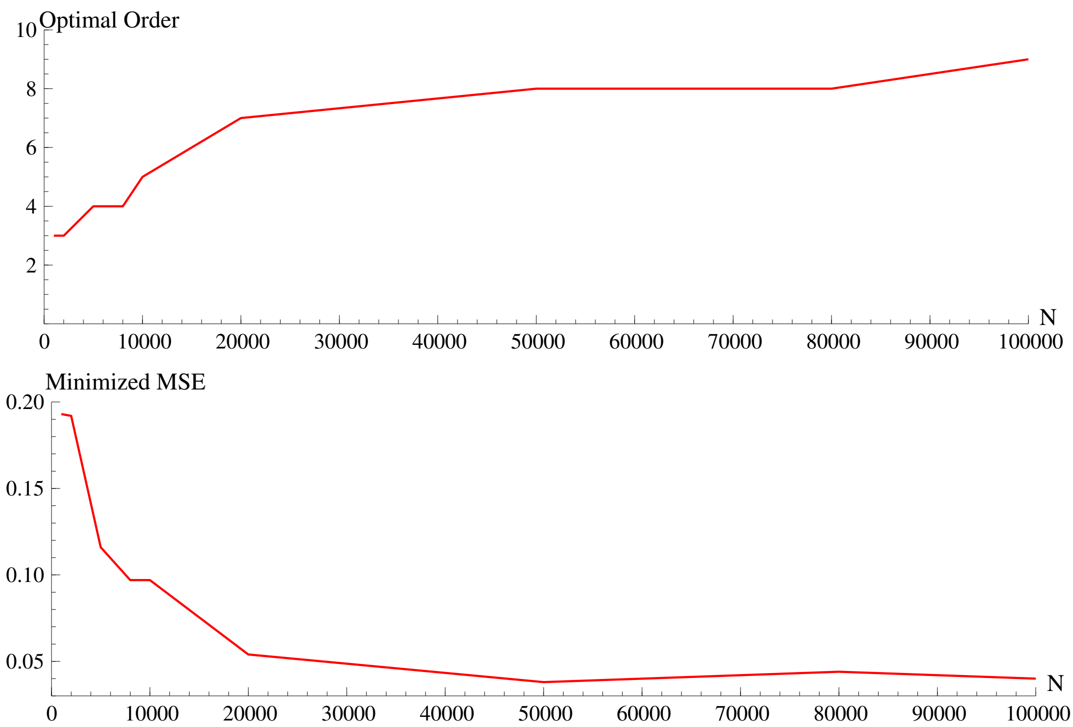

Figure 1.5: The Optimal Approximation Order

1.6.2 The Value of ρ and the Optimal Approximation Order

We now turn to investigate the relation between the value of ρ and the optimal approximation order. We set $\rho = 1$ and redo all the Monte Carlo simulations of the previous subsection. The β_2 estimates and the estimation efficiency are qualitative similar to that of the $\rho = 3$ case. we thus choose not to report them here but to summarize the MSEs in Table 1.8.

Within the CRRA utility framework, the value of ρ governs the curvature of the utility function and thus the extent of nonlinearity of the Euler equation. Intuitively, we would expect that the approximation bias is smaller when $\rho = 1$, because the Euler equations exhibit less curvature for consumers who are less risk averse. In the table, the MSEs of the 2nd-order approximated estimation are now far less than $\rho = 3$ case. This reveals the fact that ignoring higher-order consumption induces smaller approximation bias as the MSE now reduces from a magnitude of about 0.500 to 0.070.[10] We also expect that fewer approximation order is needed to achieve a minimum MSE when $\rho = 1$. These expectations can be verified by our simulation and can be seen apparently in Figure 1.6.

In the upper panel of Figure 1.6, we find that the optimal approximation order is now smaller than the $\rho = 3$ case. The required approximation order lies everywhere underneath the $\rho = 3$ line. Specifically, the optimal approximation order is now only 4 when N=10000 (which is 5 in the $\rho = 3$ case), and is 6 when N=100000 (which is 9 in the $\rho = 3$ case).

In the lower panel, the MSE of the $\rho = 1$ case exhibits similar pattern to the $\rho = 3$ case. That is, when we increase the sample size, a smaller MSE can always be achieved when the optimal approximation order is chosen. Moreover, the less curved utility function also raised the estimation performance of each sample size, as the MSE curve now lies underneath the $\rho = 3$ case.

[10] It should be stressed, however, that this does not mean that we can safely ignore the higher-order terms when $\rho = 1$. Though the approximation biases are not as large, they still yield a 100% rejection rate of the null that β_2 is equal to its theoretical value of 1.

Table 1.8: MSE of β_2 Estimates ($\rho = 1$)

Order	N=1000	N=2000	N=5000	N=8000	N=10000	N=20000	N=50000	N=80000	N=100000
2	0.088	0.082	0.076	0.075	0.075	0.074	0.074	0.074	0.074
3	0.043*	0.039*	0.032*	0.029*	0.029	0.028	0.026	0.028	0.026
4	0.052	0.051	0.039	0.036	0.026*	0.026*	0.026	0.026	0.028
5	0.072	0.056	0.043	0.042	0.039	0.038	0.024*	0.022 *	0.022
6	0.155	0.087	0.053	0.062	0.079	0.114	0.124	0.081	0.018*
7	0.241	0.123	0.062	0.067	0.085	0.121	0.140	0.103	0.066
8	0.304	0.306	0.230	0.210	0.210	0.203	0.152	0.165	0.115
9	0.396	0.326	0.288	0.302	0.244	0.222	0.154	0.172	0.166
10	0.403	0.345	0.314	0.304	0.284	0.225	0.192	0.173	0.166

Notes. Replication for 1,000 times.

* The 'optimal approximation order' with the minimum MSE.

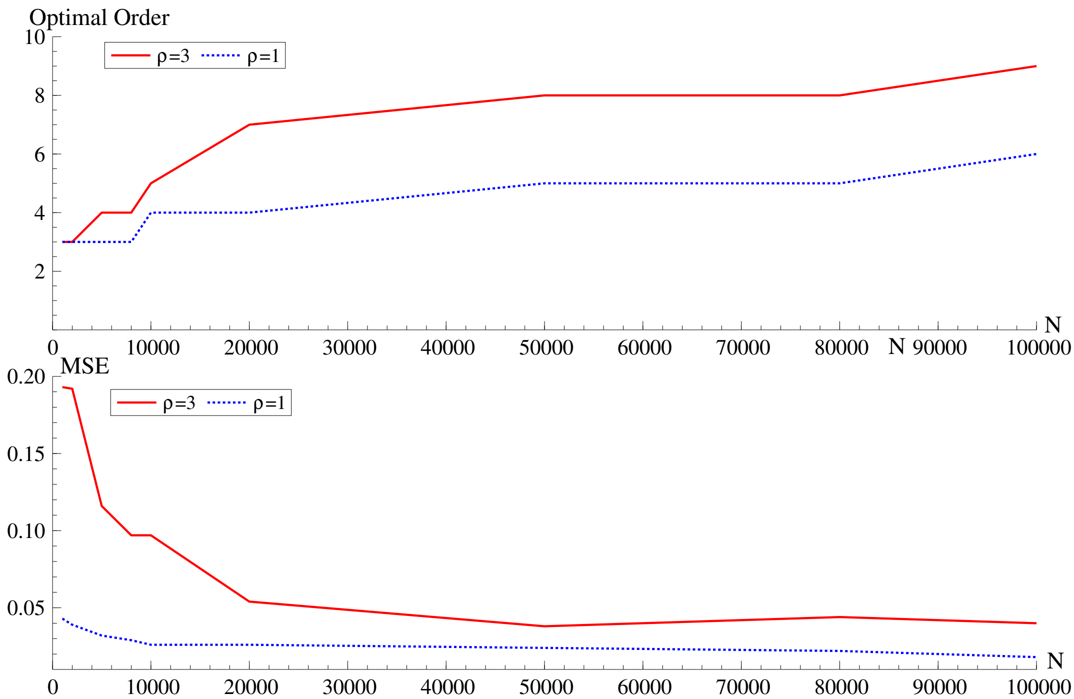

Figure 1.6: The Optimal Approximation Order of Different Values of ρ

1.6.3 Criteria of Selecting the Approximation Order

The MSE criterion serves as a good measure in selecting the optimal approximation as we are now trading off between the reduction in bias and the loss in efficiency when the higher-order moments are introduced into the model. From the previous analysis, we also realize that the optimal approximation order increases with the sample size, as well as the value of ρ. Nevertheless, the value of ρ is rarely known *ex ante* and thus the MSE criterion can not be used in empirical work. We therefore seek to find some criteria that can help selecting the approximation order.

Selecting the approximation order can be viewed as a model selection problem. Choosing among different approximation order is essentially a model selection problem among modes that encompass various consumption moments. We thus explore the usefulness of the model selection criteria that have been used so often in empirical work in our order selection problem. The performance of these model selection criteria is then evaluated with the MSE of the β_2 estimates.

1.6.3.1 Selection Based on the Goodness-of-Fit

The idea is to choose a model that best fit the data. While being intuitive, it is well known that a larger model is always preferred based on this criterion even if some variables should not have been included in the model. Modifications of the criterion then impose penalties on overparameterization to ensure parsimony. We consider three model selection criteria here:

1. Generalized adjusted R^2 ($G\bar{R}^2$):

   This criterion is proposed by Pesaran and Smith (1994), particularly for regression models estimated by the instrumental variables method. They argue that the conventional R^2 or $\bar{R}^2$ of a regression model estimated by the IV method are not guaranteed to be positive, even if the intercept term has been included in the model. More importantly, the use of R^2 or $\bar{R}^2$ do not guarantee that the true model is chosen, even asymptotically. They thus propose new goodness-of-fit measure for IV estimations that is based on prediction errors. The $G\bar{R}^2$ for model i can be written as:

$$G\bar{R}_i^2 = 1 - \frac{(\hat{e}_i'\hat{e}_i)/(N-k_i)}{S_{yy}/N},$$

where $\hat{e}_i = (I_N - \hat{X}_i(\hat{X}_i'\hat{X}_i)^{-1}\hat{X}_i)y$; $\hat{X}_i = Z_i(Z_i'Z_i)^{-1}Z_i'X_i$. N is the sample and k_i is the rank of the independent variables of model i, and $S_{yy} = \sum_{j=1}^{N}(y_j - \bar{y})^2$ is the total sum of square of the dependent variables. Note that the $\hat{e}_i$ is the residual vector from the second stage regression of the two stage least squares procedure, rather than the conventional IV residuals. The model that yields the greatest $G\bar{R}^2$ is chosen.

2. Akaike Information Criterion (AIC)

   This information criterion imposes a constant penalty on the overparameterization. It is an inconsistent model selection criterion as the probability of choosing the correct model does not approach one as the sample size goes to infinity. Despite this concern, the AIC has been a popular method of model selection in applied work. The AIC criterion takes the form:

$$AIC_i = \log(\hat{\sigma}_i^2) + 2\frac{k_i}{N},$$

   where $\hat{\sigma}_i^2 = (\hat{e}_i'\hat{e}_i)/N$. The model that yields the least AIC is chosen.

3. Bayesian Information Criterion (BIC)

   The BIC places a larger penalty on the number of parameters to be estimated than the AIC. It is therefore more parsimonious in selecting the model. The BIC criterion is

$$BIC_i = \log(\hat{\sigma}_i^2) + \log(N)\frac{k_i}{N}.$$

   The imposed penalty on the parameterization depends on the sample size N. This BIC criterion is consistent in that as the sample size goes to infinity, the probability of choosing the correct model goes to one.

1.6.3.2 Selection Based on the GMM Selection Criteria

Andrews (1999) and Andrews and Lu (2001) develop consistent model and moment selection criteria in the GMM estimation framework. Their criteria, known as the MMSC (model and moment selection criteria), are able to select the correct model and moment condition with probability equal to one asymptotically. The MMSC are based on the J test statistic that is used to test the validity of the over-identifying restrictions. Additional 'bonus terms' that reward the use of more moment conditions for a given number of parameters and the use of less parameters for a given number of moment conditions are added to the conventional J statistic.

The MMSC have many variants, here we introduce three criteria that are also introduced in Andrews and Lu (2001). These criteria are:

1. MMSC-AIC

   The MMSC-AIC is the analogue of the AIC model selection criterion. It takes the form:

   $$\text{MMSC-AIC} = J_i - 2(l_i - k_i),$$

   where J_i is the J statistic for testing over-identifying restrictions, evaluated under the specification of model i. Note that the specification of 'model i' now includes the model to be estimated as well as the moment condition used to estimate the model. The k_i is defined as before as the number of parameters to be estimated and l_i refers to the number of moment conditions under model i. The 'bonus term', $2(l_i - k_i)$, reward the use of more over-identifying restrictions at a constant rate of 2, just as the AIC criterion. The reason why this bonus term is added is that J will always increase in value when more moment conditions are added, even if these are correct restrictions. A bonus term is thus need to compensate for this. The MMSC-AIC then chooses the model that yields the minimum value of MMSC-AIC.

2. MMSC-BIC

The bonus term now depends on the sample size N, just as the BIC criterion:

$$\text{MMSC-BIC} = J_i - \log(N)(l_i - k_i).$$

The bonus of imposing the over-identification restrictions thus increases without bound as N increases. The MMSC-AIC procedure does not have this property and is thus an inconsistent model and moment selection criteria.[11]

3. MMSC-HQIC

   This criterion resembles the HQIC model selection criterion of Hannan and Quinn (1979). The MMSC-HQIC is defined by:

$$\text{MMSC-HQIC} = J_i - Q\log(\log(N))(l_i - k_i),$$

   for some $Q > 2$. In the Monte Carlo experiments that follow, we follow Andrews and Lu (2001) in setting $Q = 2.1$ throughout this chapter.

1.6.4 Performance of the Model Selection Criteria

We now evaluate the performance of the 6 model selection criteria in selecting the appropriate approximation order to the consumption Euler equation. It should be noted, however, that the environment of our analysis is different from that of the Andrews (1999) and Andrews and Lu (2001). In selecting the approximation order, there is actually no 'correct' model, because the correct approximation order is infinity when the utility function is set to be a CRRA one. Yet as we mentioned earlier, infinitely approximating the Euler equation is infeasible because the efficiency loss from the multi-collinearity problem would be too huge to be tolerable.

Despite this difference, we are still interested in investigating the usefulness of these criteria in selecting the *appropriate* approximation order. Monte Carlo studies with different sample sizes are thus conducted. Consistent with the previous studies, the simulations are replicated for 1,000 times. The MSEs of the β_2 estimates chosen by different criteria are then compared to evaluate the performance of each criterion Before the comparison

[11] Interested readers may consult Andrews (1999) and Andrews and Lu (2001) for detailed proof.

of the MSE, however, we now present first the information on the distribution of the approximation order selected by each model selection criterion. This will facilitate the analysis of the MSE comparisons.

Table 1.9 reports the average selected order from 1,000 replications.[12] In the parentheses, the standard error of the selected order are also reported. For the average selected order, if we compare among the 3 criteria based on the goodness-of-fit, we find that the BIC criterion is the most parsimonious, and that the $G\bar{R}^2$ tends to select the most order. This is not surprising as the BIC criterion imposes the most penalty on overparameterization, and the least penalty is imposed in the $G\bar{R}^2$. We also find that the selected order increases with the sample size, regardless of the chosen criterion. If we compare the order selected here and the 'optimal approximation order' in Table 1.7, we can conclude that the $G\bar{R}^2$ and the AIC criterion tends to overfit. On the other hand, the BIC criterion seems to yield mean selected order that is fairly close to the optimal approximation order that yields the minimum MSE in table Table 1.7.

Turning now to the MMSC procedures proposed by Andrews and Lu (2001), the first observation is that the order selected by these 3 criteria also increases with the sample size. The MMSE-BIC seems to be most parsimonious, and the MMSE-AIC tends to select the most order. The order selected by MMSE-HQIC then lies between these two criteria. This is also expectable, as the MMSC-BIC places most reward on the $l-k$ (the number of over-identifying restrictions in excess of the number of parameters). Since the number of moment conditions does not vary with the selected order (specifically, $l_i = 16$ for all the approximation order), this is equivalent to imposing the most penalty on the inclusion of additional order. The order selected by the MMSC-BIC thus turns out to be most parsimonious. For the sample size N between 1000 and 100000, $Q\log(\log(N))$ lies between 2 and $\log(N)$), the MMSC-HQIC thus select the order that lies between that selected by the MMSC-BIC and MMSC-AIC.

For the robustness of the approximation order selected by each criterion, Table 1.9

[12] For the 16 instrumental variables (moment conditions) we are using, the Euler equation can only be approximated up to the 16th order. We thus set 16 as the upper bound of the approximation order in these model selection criteria.

Table 1.9: Approximation Order Selected by Different Criteria

Criteria	N=1000	N=2000	N=5000	N=8000	N=10000	N=20000	N=50000	N=80000	N=100000
$G\bar{R}^2$	8.733	9.258	9.935	10.196	10.119	10.345	10.220	10.460	10.465
	(1.519)	(1.608)	(1.554)	(1.624)	(1.773)	(1.992)	(2.019)	(2.014)	(2.015)
AIC	7.014	7.638	8.856	9.195	9.282	9.563	9.63	9.889	9.900
	(2.184)	(2.045)	(1.748)	(1.764)	(1.808)	(1.864)	(1.858)	(1.816)	(1.809)
BIC	3.044	3.262	4.334	5.233	5.821	7.168	8.037	8.445	8.561
	(1.240)	(1.427)	(1.869)	(1.997)	(1.953)	(1.438)	(1.264)	(1.265)	(1.263)
MMSC-AIC	4.566	4.985	5.435	5.464	5.363	5.504	5.655	5.765	5.835
	(1.817)	(1.780)	(1.485)	(1.430)	(1.385)	(1.236)	(1.084)	(1.008)	(1.004)
MMSC-BIC	2.253	2.327	2.742	3.228	3.349	3.814	4.031	4.606	4.808
	(0.652)	(0.762)	(1.031)	(1.256)	(1.249)	(1.256)	(1.212)	(1.010)	(0.945)
MMSC-HQIC	2.884	3.176	4.033	4.380	4.440	4.531	4.996	5.221	5.349
	(1.263)	(1.406)	(1.512)	(1.505)	(1.444)	(1.349)	(1.064)	(0.953)	(0.925)

Notes. Figures reported in the table are the average approximation order selected by each criteria. Standard deviation of the selected order of the 1,000 replications in parentheses. The upper bound of the approximation order is fixed at 16.

reveals that the criteria based on the goodness-of-fit tends to yield greater variations in the selected approximation order, especially the $G\bar{R}^2$ and the AIC. The problem becomes more prominent when the sample size increases. As is shown in the table, this increase in the sample size does not drive down the variation of the selected order. A standard error of the magnitude around 2 is quite large compared with that of the selected order around 1 for the MMSC procedures. This would lead to devastating effect on the MSE of those criteria based on the goodness-of-fit, which would be discussed later in this section.

The distributions of the selected order are depicted in Figure 1.7. In the figure, the sample size is set at $N = 100,000$. The orders selected by the $G\bar{R}^2$ and the AIC are quite widespread between the 7th and the 16th order, while the MMSC procedures yield more condensed order distributions between the 3rd and the 7th order.

The post-selection estimation results are summarized in Table 1.10. For each criterion considered, we estimate β_2 with the approximation order chosen by the criterion. The average estimate as well as the standard deviation of the 1,000 replications are reported

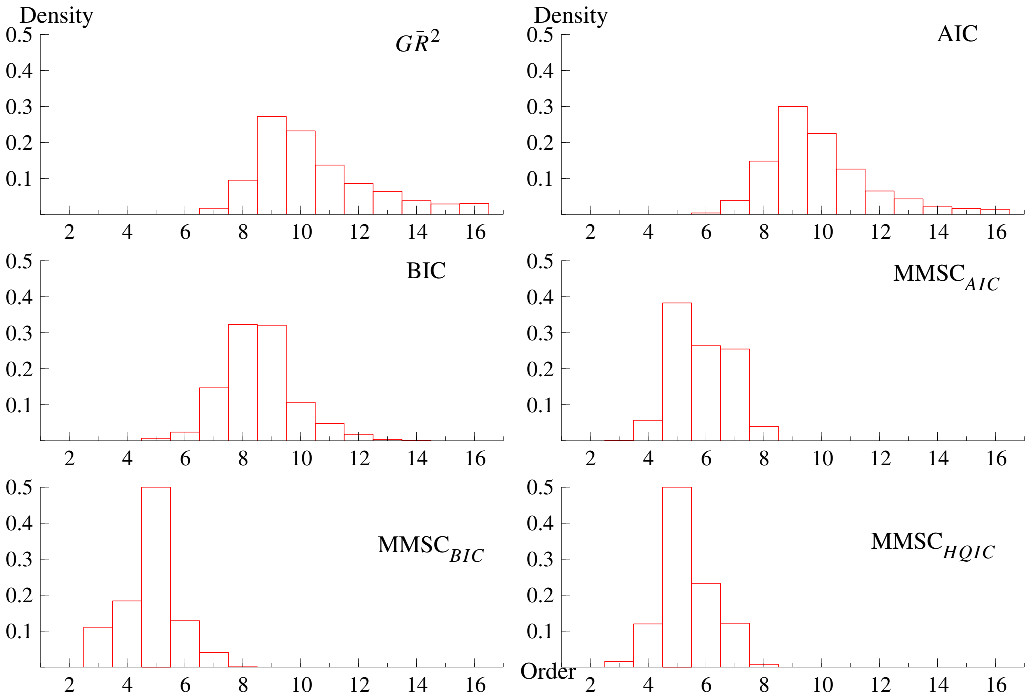

Figure 1.7: The Distribution of the Selected Order (N=100000)

in the table. For the small sample size such as $N = 1000$, none of the selection criteria can provide mean estimate that is close enough to the true value of 2. The variations of the estimates are also fairly large, especially for estimates from the order selected by the $G\bar{R}^2$ and the AIC. The reason is quite clear from Table1.6. When $N = 1000$, for each of the approximation order under consideration, the β_2 estimates do not approach the true value and tend to vary in a wide range. This poor prior-selection estimation result and the great variations in the selected order thus lead to the poor small sample post-selection estimation performance.

When we increase the sample size, however, the improvement in the estimation performance is quite evident. Take the MMSC-AIC as an example, the bias can be reduced from 0.583 (2-1.417) to 0.007 (2-1.993), and the standard deviation can be reduced from 0.624 to 0.313, when we increase the sample size from 1000 to 100000. The $G\bar{R}^2$ and the AIC criteria appear to be relatively inefficient, the reason is that these two criteria tends to overfit the model. As we mentioned earlier, selecting too many orders would lead to multi-collinearity problem, and thus great loss in efficiency.

Table 1.10: β_2 Estimates Selected by Different Criteria

Criteria	N=1000	N=2000	N=5000	N=8000	N=10000	N=20000	N=50000	N=80000	N=100000
$G\bar{R}^2$	1.100	1.219	1.339	1.434	1.519	1.602	1.805	1.899	1.900
	(1.317)	(1.192)	(1.135)	(1.098)	(0.989)	(0.881)	(0.740)	(0.570)	(0.577)
AIC	1.219	1.366	1.456	1.548	1.590	1.660	1.815	1.897	1.902
	(1.008)	(0.902)	(1.060)	(0.877)	(0.957)	(0.852)	(0.733)	(0.581)	(0.577)
BIC	1.461	1.524	1.694	1.807	1.875	1.954	1.905	1.934	1.925
	(0.369)	(0.927)	(0.339)	(0.400)	(0.379)	(0.420)	(0.519)	(0.442)	(0.381)
MMSC-AIC	1.417	1.596	1.765	1.829	1.823	1.895	1.954	1.982	1.993
	(0.624)	(0.541)	(0.436)	(0.335)	(0.307)	(0.277)	(0.310)	(0.329)	(0.313)
MMSC-BIC	1.319	1.370	1.495	1.586	1.620	1.710	1.743	1.853	1.869
	(0.295)	(0.272)	(0.339)	(0.348)	(0.354)	(0.339)	(0.327)	(0.332)	(0.339)
MMSC-HQIC	1.355	1.466	1.644	1.730	1.738	1.799	1.902	1.949	1.952
	(0.430)	(0.393)	(0.383)	(0.359)	(0.348)	(0.321)	(0.313)	(0.346)	(0.354)

Notes. Figures reported in the table are the average β_2 estimates selected by each criteria. Standard deviation of the 1,000 replications in parentheses.

The MMSC-BIC, on the other hand, can yield relatively efficient post-selection β_2 estimates. But they are subject to greater biases. The reason is that the MMSC-BIC procedure is too parsimonous in selection the approximation order. Insufficient approximation order thus leads to mean estimates that are more biased away from the true value.

To evaluate the overall performance of each model selection criterion, the MSEs of the β_2 estimates are summarized in Table1.11 and Figure1.8. As what can be expected, the MSE declines with the increase of sample size, regardless with the chosen criterion. Nevertheless, for a given sample size, especially when the sample size is small, the performance of these various criteria may be quite different. Obviously, the $G\bar{R}^2$ and the AIC yield the worst estimation performances, implying that the multi-collinearity problem from overfitting the model might result in devastating effect on the estimation performance.

The performance of other criteria appears to be relatively close to each other, especially when the sample size is large. But if we are to choose a criterion among them, the MMSC-AIC would be our first choice. For a sample size greater than 5,000, it yields the minimum

Table 1.11: MSE of β_2 Estimates Selected by Different Criteria

Criteria	N=1000	N=2000	N=5000	N=8000	N=10000	N=20000	N=50000	N=80000	N=100000
$G\bar{R}^2$	2.545	2.031	1.726	1.525	1.207	0.934	0.586	0.335	0.343
AIC	1.626	1.216	1.421	0.975	1.084	0.841	0.572	0.348	0.343
BIC	0.427	0.313	0.208	0.197	0.160	0.178	0.278	0.199	0.151
MMSC-AIC	0.729	0.456	0.245	0.141	0.125	0.088	0.099	0.108	0.098
MMSC-BIC	0.551	0.470	0.370	0.293	0.270	0.199	0.173	0.132	0.132
MMSC-HQIC	0.601	0.440	0.274	0.202	0.190	0.143	0.108	0.123	0.127

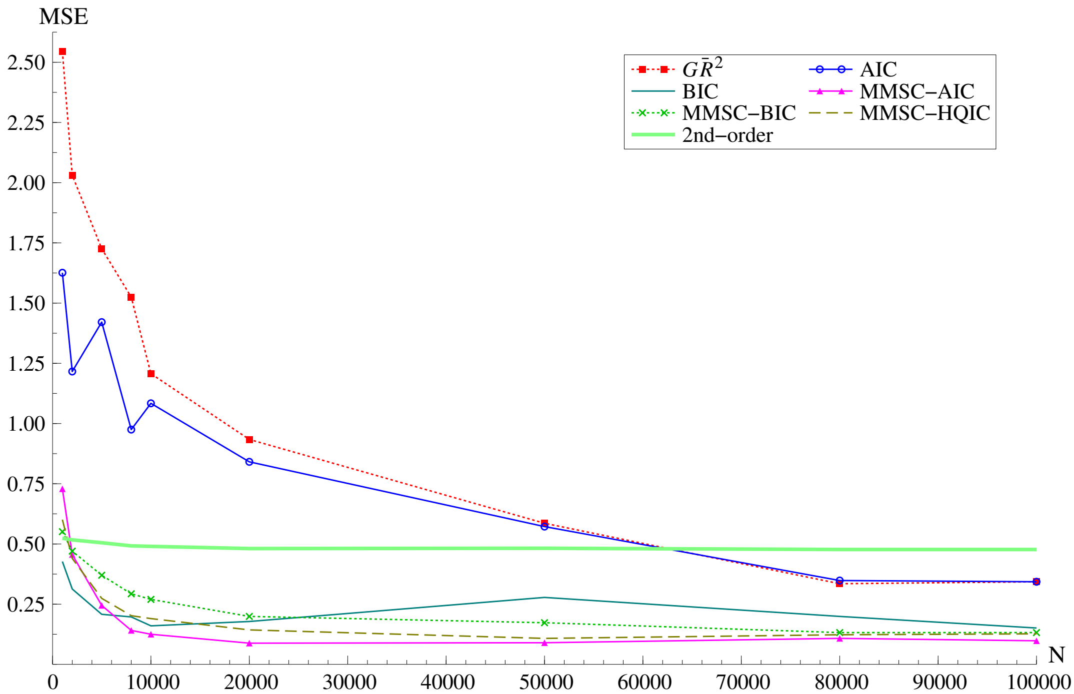

Figure 1.8: MSE of the β_2 Estimates from Different Model Selection Criteria

MSE among the criteria introduced in this chapter.

We also depict the MSE of the β_2 estimates from the 2nd-order approximation to the consumption Euler equation, which is labeled '2nd-order' in Figure1.8. As is evident from the table, for sample size greater than 1000, a smaller MSE of β_2 estimates can always be achieved with the approximation order suggested by the MMSC or BIC.

To conclude this section, we find that improper linear approximation may lead to very poor performance in uncovering the structure parameters. Nevertheless, for a reasonable sample size that is available in most data sets such as the PSID, estimates of the the structure parameters can be fairly close to their true values, if the proper approximation order is chosen. This approximation order is directly linked to the sample size and the value of ρ, and has to be carefully chosen in estimation. Our Monte Carlo experiment reveals that with the employment of the model selection criteria suggested here, there is ample room for the post-selection estimators to outperform the conventional second-order approximation that has been so popular in empirical work.

1.7 Conclusion

This chapter investigates the possibility of reducing approximation bias by implementing higher-order approximations of the consumption Euler equation. Our result is in sharp contrast to the arguments made by Attanasio and Low (2004), and Alan and Browning (2003). They argue that trying to estimate the price elasticity (the EIS) without the intertemporal variations in price is destined to fail. Nevertheless, without variations in the interest rate, we find that the inclusion of higher-order consumption moments does provide improvement over the second-order approximated version of the Euler equation that has been a popular choice of model specification for purposes of estimation. The estimates of the structural parameters can be fairly close to their true value if the Euler equation is approximated to the appropriate order.

This reduction in approximation bias, however, is not achieved without any price being paid. We find that there is a trade-off between the reduction of approximation bias and the the loss in efficiency when higher-order moments are introduced into the model. We thus seek to find the 'optimal approximation order' that can yield the minimum MSE of the estimates of the structure parameters. This optimum order is found to be directly linked to the value of ρ and the the sample size. Moreover, using MSE criteria, a second-order approximated version of the Euler equation is never the best approach to uncover the structural parameters. A substantially better-off situation could be achieved simply by allowing for higher-order moments in the consumption growth regression.

Apart from pointing out the problems associated with second-order approximation of the consumption Euler equation, we have demonstrated in this chapter a practical alternative for estimating structural parameters that is subject to less approximation bias. Yet we also find that the efficiency loss when higher-order moments are included may be huge. This efficiency loss results from the difficulty in identifying independent variations in higher-order moments by sets of linear instruments used to identify that of the second-order moment. Finding quality instruments that can better explain the higher-order moments thus may provide better performance in estimation. We seek to identify these instruments from the nonlinear relationship between higher-order moments in the next chapter.